Midterm exam Statistical Methods – Solutions

17 November 2015

1.

a) True. The mean is given by $(x_1 + \ldots + x_n)/n$ so it always changes with the change of any value x_i.

b) False. Results that are very easy to get are used in convenience sampling.

c) False. Bar chart can be used only for qualitative data, at the nominal or ordinal level of measurement.

d) False. Download speeds are at the ratio level of measurement, since there is a natural zero starting point.

NB. *Other examples/arguments are of course also possible.*

2.

a) From the boxplot we can infer that the data has a median around -1. The 0.25-quantile of the sample is around -1.2 and the 0.75-quantile is around -0.6. Therefore, the interquartile range is approximately $-0.6 - (-1.2) = 0.6$, which describes the spread of the data. The boxplot indicates that the data is symmetric, with three outliers on the left-hand side.

b) The area of a bar in the histogram is equal to the relative frequency of the data in the interval below the bar.

c) When we draw a (imaginary) straight line through the middle of the QQ plot, we see that the left part of the plot is below this straight line. Therefore, the left tail of the underlying distribution of the data is heavier than that of a normal distribution.

3.

a) We first need to find the probability $P(\text{Windows})$ that a randomly selected user has a Windows phone. Note that $P(\text{Windows}) = 1 - P(\text{Android}) - P(\text{iPhone}) = 1 - 0.2 - 0.65 = 0.15$. Since the app is popular, we follow the small sample rule, i.e., we treat selections as independent events. Therefore, by the multiplication rule for independent events we get

$$P(\text{two randomly selected users have a Windows phone}) \\ = P(\text{Windows}) \cdot P(\text{Windows}) = 0.15 \cdot 0.15 = 0.0225.$$

b) Using the law of total probability we get

$$P(\text{fee}) \\ = P(\text{fee}|\text{Android}) \cdot P(\text{Android}) + P(\text{fee}|\text{iPhone}) \cdot P(\text{iPhone}) + P(\text{fee}|\text{Windows}) \cdot P(\text{Windows}) \\ = 0.1 \cdot 0.2 + 0.6 \cdot 0.65 + 0.2 \cdot 0.15 = 0.02 + 0.39 + 0.03 = 0.44.$$

c) We are interested in $P(\text{iPhone}|\text{fee})$. By definition of conditional probability

$$P(\text{iPhone}|\text{fee}) = \frac{P(\text{iPhone and fee})}{P(\text{fee})}.$$

Using the multiplication rule in the numerator and the answer from part b) in the denominator we get

$$P(\text{iPhone}|\text{fee}) = \frac{P(\text{fee}|\text{iPhone}) \cdot P(\text{iPhone})}{P(\text{fee})} = \frac{0.6 \cdot 0.65}{0.44} \approx 0.886.$$

The same answer can be obtained by using Bayes' theorem.

4.

a) The sample space Ω is $\{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$. All outcomes are equally likely, since the coin is fair, thus $P(\omega) = 1/8$ for all ω in Ω.

b) Two events A and B are independent if $P(A \cap B) = P(A) \cdot P(B)$.
We have $A = \{HHH, HHT, THH, THT\}$, so $P(A) = 1/2$. Similarly, $B = \{HTT, THT, TTH\}$, so $P(B) = 3/8$. Finally, $A \cap B = \{THT\}$ and $P(A \cap B) = 1/8$. However, $P(A) \cdot P(B) = 1/2 \cdot 3/8 = 3/16 \neq 1/8 = P(A \cap B)$. We conclude that A and B are not independent.

c) We find what Alice earns for each outcome of the experiment, and get the following table:

x	$\{X = x\}$	$P(\{X = x\})$
6	$\{HHH\}$	$1/8$
4	$\{HTH, THH\}$	$1/8 + 1/8 = 1/4$
2	$\{TTH\}$	$1/8$
0	$\{HHT, HTT, THT, TTT\}$	$1/8 + 1/8 + 1/8 + 1/8 = 1/2$

d) By definition of expectation we get

$$E(X) = \sum_x x \cdot P(X = x) = 6 \cdot \frac{1}{8} + 4 \cdot \frac{1}{4} + 2 \cdot \frac{1}{8} + 0 \cdot \frac{1}{2} = 2.$$

5.

a) Printing time of a single one A4 page is assumed to be normally distributed with mean $\mu = 10$ and standard deviation $\sigma = 1.5$. Therefore, the z-score of 9 seconds is $z = \frac{9-10}{1.5} \approx -0.67$. We use Table 2 and see that $P(Z \leq -0.67) = 0.2514$, where Z denotes the standard normal distribution. Hence, the probability that a single page will be printed in less than 9.00 seconds is 0.2514 (or 25.14%).

b) The mean printing time of $n = 25$ randomly selected pages is normally distributed with mean $\mu = 10$ and standard deviation $\sigma = 1.5/\sqrt{25} = 1.5/5 = 0.3$, by the Central Limit Theorem. Therefore, the z-score of 10.66 seconds is $z = \frac{10.66-10}{0.3} = 2.2$. We use Table 2 again and see that $P(Z \leq 2.2) = 0.9861$, so $P(Z \geq 2.2) = 1 - 0.9861 = 0.0139$. Hence, the probability that the mean printing time of 25 pages is more than 10.66 seconds is 0.0139 (or 1.39%).